An Approach to Data Extraction and Visualisation for Wireless Sensor Networks

Mohammad Hammoudeh, Robert Newman, Sarah Mount
School of Computing and IT
University of Wolverhampton
Wolverhampton, UK
Email: {m.h.h, r.newman, s.mount}@wlv.ac.uk

***Abstract*—Ever since Descartes introduced planar coordinate systems, visual representations of data have become a widely accepted way of describing scientific phenomena. Modern advances in measurement and instrumentation have required increasingly sophisticated visual representations, to ensure that scientists can quickly and accurately interpret increasingly complex data. Most recently, wireless sensor networks (WSNs) have emerged as a technology which is capable of collecting a vast amount of data over space and time. The sheer volume of the data makes it difficult to be interpreted by humans into meaningful insights. This presents a number of challenges for developers of visualisation techniques which seek to "map" the data sensed by a network. Visualisation techniques helps to turn large amounts of raw data into credible visual information such as graphs, charts, or maps, that can assist in understanding of the meaning of that data. In this paper we propose a *map* as a suitable data visualisation and extraction tool. We aim to develop an in-network distributed information extraction and visualisation service. Such a service would greatly simplify the production of higher-level information-rich representations suitable for informing other network services and the delivery of field information visualisation.**

I. Introduction

The main objective of a wireless sensor network (WSN) is to provide users with access to the information of interest from data gathered by spatially distributed sensors. In real-world applications, WSNs are often deployed in a high density to ensure a full coverage of the monitored phenomena. These networks are expected to generate an excessive amount of data. As the sensor network scales in size, so does the amount of sensed data which is required to be collected by the network, processed, and presented to the user. The data produced and the form in which they are structured varies widely. Scientists need tools to reconcile these differences. Human interpretation of this data will require extensive use of visualisation tools. Using scientific visualisation techniques can help to make meaningful visualisation possible in many aspects of data understanding and analysis. Data visualisation is becoming increasingly important in WSNs as it enables end-users to best utilise the data collected by the sensor network.

In this paper we propose a map as a suitable visualisation and data extraction tool for WSNs. A map is a visual representation of an area, although most commonly used to depict geography, maps may represent any space, real or imagined, without regard to context or scale such as weather data distribution mapping [1]. It could be overlaid over a geographic map to enable observer to view the data for a specific area.

This paper is organised as follows. Section 2 discusses the related work. Section 3 presents the characteristics of the sense data. In Section 4, we discuss visualisation challenges in WSNs. In Section 5 we introduce the benefits of visualisation of sense data. In Section 6 we discuss the advantages of map data format. We present the implementation of the mapping services in Section 7. We also evaluate the performance of the proposed mapping service in Section 8. And we conclude the work in Section 9.

II. Related Work

Within the WSN field, mapping applications found in the literature are ultimately concerned with the problem of mapping measurements onto a model of the environment. Hellerstein et al. [2] propose to construct isobar maps in sensor networks. They show how in-network merging of isobars could help reduce the amount of communication. Furthermore, [3] proposes an efficient data-collection scheme, and the building of contour maps, for event monitoring and network-wide diagnosis, in centralised networks. Solutions such as Distributed Mapping have been proposed to the general mapping domain [2]. However, many solutions are limited to particular applications and constrained with unreliable assumptions. The grid alignment in [2], for example, is one such assumption.

In the wider literature, mapping was sought as a useful tool in respect to network diagnosis and monitoring [3], power management [4], and jammed-area detections [5]. For instance, contour maps were found to be an effective solution to the pattern matching problem that works for limited resources networks [6]. As opposed to resolving these types of isolated concerns, in the work proposed here the WSN is expected not only to produce map type responses to queries but also to make use of the data supporting the maps for more effective routing, further intelligent data aggregation and information extraction, power scheduling and other network processes.

These are examples of specific instances of the mapping problem and, as such, motivate the development of a generic distributed, in-network mapping framework, furthering the area of research by moving beyond the limitations of the centralised approaches.

III. Characteristics of Sense Data

Data acquired from a WSN is imperfect in nature. This imperfect nature of data is due to physical constraints on node deployment and data collection, noisy environment, device measurement errors, among other factors. Moreover, data collected by different sensors may have various qualities depending on physical characteristics such as distance from sensed phenomena, node modality, or noise model of individual sensors [7].

In densely deployed WSNs, sensor readings are usually highly correlated in the space domain [8]. Additionally, the nature of the physical phenomenon contains the temporal correlation between each sensor node successive readings. Data gathered from a WSN is often characterised by its significant redundancy. Many sensor networks are densely deployed with high node redundancy to deal with node failure based connectivity and coverage problems. However, dense deployment causes neighbouring sensor nodes to have highly overlapping sensing regions. Consequently, it is likely that multiple nodes often detect and communicate data packets about common phenomena. This is due to the fact that each node observes the physical region of overlap independent of its neighbours. Data aggregation techniques aim at reducing redundant data transmissions. Although data aggregation results in fewer transmissions, however, they introduce considerable amount of delays on delivering the data to its final destination. Data from nearer sources may have to be held back at an intermediate node in order to be aggregated with data coming from sources that are farther away. In addition, within such networks, delays may be caused by hop-by-hop retransmission, scheduled data communication, queueing delays, propagation through the environment, and other factors.

In many WSNs application areas, such as medical or surveillance applications, the accuracy of acquired data is often crucial. However, sensed data is often inaccurate and erroneous [9]. This inaccuracy may be a result of faulty sensor readings, internal errors in sensor nodes, network delays, among other reason. The deployment of a larger number of sensor nodes provides potential for greater accuracy in the information gathered. The ability to effectively increase the sensing quality without necessarily increasing data transmissions will increase the reliability of the information for the end user application. Some schemes such as [10], trade data accuracy for energy efficiency, which typically increase with the amount of data transmissions. Data aggregation also increases the level of gathered data accuracy and exploits data redundancy to compensate node failures [10], [9]. Depending on the accuracy bounds required for a specific application, a node may need to communicate some of its information to the sink that is incorporated in the model so that the accuracy bounds are met.

Another characteristic of sampled sense data is the distribution of sampled source data. The distribution of data is usually specified in terms of location and pattern [11]. The location is specified in terms of Cartesian coordinates (x, y, z), while the pattern falls into one of two categories: regular, when data is gathered from sensing nodes which are deployed on a grid; irregular, when sensing nodes are deployed randomly, many sensor network data belongs to the irregular patterns category. It is sometimes necessary to know the data density besides their distribution. The network density is defined as the number of nodes per unit area.

IV. Why visualisation of Sense Data is Challenging

Visualisation of sensor network data is challenging. The large amount of data collected from deployed sensor networks arrives in "bursty" mode. This makes it difficult to process all the data in a timely manner so that it can be used as an input to visualisation systems [12]. Furthermore, most scientific visualisation techniques require data to include connectivity information, which is not provided by a scattered data set. Hence, highly efficient visualisation schemes operating directly on raw scattered data are necessary [13].

In large-scale WSN it is a non-trivial task to visualise the observed phenomena given the problems of sparse, inaccurate, high density, and irregularly distributed data, in addition to the limited physical resources. Therefore, we aim to define methods suitable for real-time visualisation and analysis suitable for sensor network data. This includes the specification of a modular scattered data interpolation package for implementing suitable interactive viewers for time-varying data.

Practically, a point measured on a surface represents the environment conditions over an area of a certain size, hence it should be possible to generate a complete view of the observation field of a WSN using a discrete set of observation points. Depending on the desired degree of accuracy and fidelity, the sampling interval over an unknown surface is defined. The problem is to define how to adequately represent the observed phenomena by a limited number of elevation points, that is, what sampling interval to use with an unknown surface? To select between different sampling strategies when data is acquired from a sensor network many factors must be considered including: application, level of accuracy, the nature of the sampled data, the nature of monitored environment, nodes distribution, and density. Data acquisition strategy is of high importance in sensor networks and is directly related to routing and single node capabilities. When sampled data is to visualise a certain observed phenomena results are good as the sample [11].

In WSNs, it is increasingly becoming important to have a live picture of the changing environmental variables. This can be achieved if data about the phenomenon is sampled at a suitable rate. In live data representation applications, where a client is interested in the current picture of the environment, it is important to collect readings from the network at a rate close as possible to the rate of change of the monitored environmental variable. In other applications, such as temperature and pressure monitoring, when the measured phenomena is changing linearly over time and space, the sampling rate can be reduced so that node resources are used effectively.

V. Benefits of Sense Data visualisation

Data visualisation in WSNs has the ability to bridge the gap between the physical and logical worlds, by using the gathered information from the physical world and communicating that information to the end-user in compact and often easy to understand way. Data visualisation helps to deal with this flood of information, integrating the human in the data analysis process. The main advantages of the application of visualisation techniques in wireless sensor networks are:

1) Visualisation could help in managing the huge amount of data coming from a sensor network. Visual data exploration can easily deal with large, highly non-homogeneous and noisy amount of data.
2) Maximisation of useful information return. Visualisation is a fundamental tool to communicate information in a compact and easy to understand way. It allows the user to gain insight into the data, drawing conclusions and directly interacting with the data. Visualisation not only helps to answer questions that user has, but it elicits questions that he did not even think of before.
3) Interactive visualisations benefit from dynamic queries which are a valuable tool to explore data.
4) More reliable information than possible from individual sources. Although individual devices have limited resources, the true value of the sensor network systems comes from the emergent behaviour that arises when data from many places in the system is combined into a meaningful presentation [14]. The bandwidth of data transferred in a picture is much bigger than having a human look at log files or textual data.
5) Detection of higher-order relationships between different sensors. Relationships become apparent. Sometimes they are completely hidden without visualisation.
6) More efficient data and information representation. Visualisation reduces analysis and response times. Going through thousands of line of points data is slower than looking at a few graphs of the same data. It is a valuable tool to communicate information in a compact and often easy to understand way.
7) Visual data examination does not require deep understanding of complex mathematical or statistical algorithms. Visualisation techniques provide a qualitative overview useful for further quantitative analysis [15]. It definitely reduces analysis and response times.

VI. Sense Data Visualisation: Maps

Visual formats, such as maps, can be easily understood by people possibly from different communities, thus allow them to derive conclusions based on substantial understanding of the available data. This understanding gained from maps fulfils the ultimate goal of sensor network deployments which is not only to gather the data from the spatially distributed sensor nodes, but also convey and translate the data for scientists to analyse and study. Visualisation of data collected from a sensor network in a map format is one way to display the distribution of the data attributes in a real-world map in ways that would be intuitive and easy to talk and reason about. A map is intuitive and easy to understand as it provides an interface for visualising dynamic data from sensor net. It provides a higher-level information-rich representation which was found suitable for informing other network services and the delivery of field information visualisation. This information-rich representation satisfies the various requirements for the sensor network system end users. The map gives a low cost interface used to target queries for generating detailed maps from a subset of the sensors in the network.

Maps are effective to understand spatial distribution of environmental features, since humans can use their natural interpretation capabilities to understand colours, patterns, and spatial relevance. The human interpretation capabilities suggest the importance of expression methods such as how to represent spatial data on a map. Maps can be either static or dynamic and allow data representation on two-dimensional or tree-dimensional space. They allow the user to infer the actual sizes of and distance between objects. The unique visualisation and analysis benefits offered by maps make them more visually communicative, they imply the distributions and states; provide information about spatial patterns; and imply the association of diverse phenomena. The users can zoom in or zoom out respectively meaning showing more or less details. Finally, representations such as maps allow to extract information that can not be obtained by looking at sensor readings separately and are more efficient to compute in both time and energy. For instance, maps may capture trends or correlations among sense data and missing data, where there is no operating sensor, can be interpolated using these spatial and temporal correlations among sensor readings.

VII. Mapping Services for WSNs

Leading directly from Section VI which shows the usefulness of visualising sense data in a map format, we investigate the development of methods for map construction and maintenance services within a WSN, focusing on service cost, complexity, computational load, storage, communication requirements, robustness to packet loss, nodes failures, and network density.

We propose a new network service: map generation. Map generation is essentially a problem of interpolation from sparse and irregular points. This service allows the production of maps of arbitrary level of detail upon requests injected into the network by the user, or pre-programmed as responses to network events. The service should be entirely based on in-network processing and would be applied to flat, computationally homogeneous networks. Given a set of known data points representing the nodes' perception of a given measurable parameter of the phenomenon, what is the most likely complete and continuous map of that parameter? In the work proposed here the WSN is expected not only to produce map type responses to queries but also to make use of the data supporting the maps for more effective routing, further intelligent data aggregation and information extraction, power

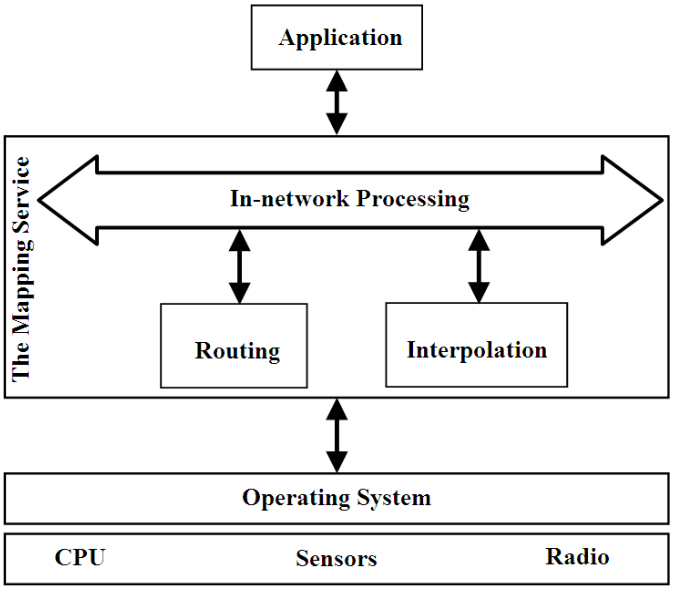

Figure 1. Architecture of the distributed in-network mapping service.

scheduling and other network processes. Just as clustering, routing and aggregation allow for more sophisticated and efficient use of the network resources, a mapping service would support other network services and make many more applications possible with little extra effort.

The proposed implementation of the distributed mapping service contains four modules as seen in Figure 1: *Application, Interpolation, In-network Processing* and *Routing*.

The Routing module is an essential module responsible for data communication. Routing proved to be a key issue as the research developed. In this paper, we use a hierarchical routing algorithm called MuMHR [16]. The defined routing procedure builds the hierarchy and establishes the path between sensing nodes and their respective cluster-head to enable data transmissions. MuMHR is an improvement over LEACH [17]. It relaxes some of the assumptions made by LEACH such as the single hop communication. The main objective of MuMHR protocol is to provide substantially energy-efficient and robust communication. The energy efficiency is achieved by load balancing at two levels: (1) at the network level, which involves traffic multiplexing over multiple paths; (2) at the cluster level, introducing rotation of the cluster-heads every given interval of time. This prevents energy depletion resulting from constantly using the same path for transmission or particular nodes being cluster-heads for a long duration. The multi-path feature is not only used for load balancing but also when path failures occur. When a path fails, an alternative path can be immediately used which allows the protocol to dynamically adapt to failures without delays or degradation in the quality of service. At the cluster set-up time, one or more nodes are chosen as cluster-head backup node(s). Backup cluster-head node substitute for the cluster-head in some failure cases or when the current cluster-head decides to reduce its participation in the protocol if its energy level approaches a certain threshold value. For instance, if the current cluster-head decides to hand its role to the backup node, it notifies the respective node and forwards to it necessary information, such as the backup nodes list, to avoid a complete cluster set-up phase.

The role of In-network Processing module is to process raw data received from various cluster-head nodes in the network. It applies filtering on all the received data to reduce redundancy resulting from overlapping cluster coverage. Moreover, the In-network Processing module manages incremental update messages and merges them into single transaction. Also, it defines two interfaces for the Interpolation and Application modules through which it provides access to the cached data in a suitable format.

All mapping applications use the Interpolation module as a building block to generate maps. The Interpolation module provides access to the In-network Processing module to obtain the available mapping or update data. In this paper, we use Shepard interpolation algorithm [18]. Shepard interpolation is simple and intuitive. It is suitable for large-scale wireless sensor networks because it reduces communication overhead by only considering data points which are significant for the interpolation results. Shepard defines a continuous function where the weighted average of data is inversely proportional to the distance from the interpolated location. This method exploits the intuitive sense that things that are close to each other are more likely to be similar. Shepard's expression for globally modelling a surface is:

$$f_1(P) = \begin{cases} \sum_{i=1}^{N} (d_i)^{-u} z_i & \text{if } d_i \neq 0 \text{ for all } D_i (u > 0) \\ z_i & \text{if } d_i = 0 \text{ for some } D_i \end{cases} \tag{1}$$

where d_i is the standard distance metric from an interpolation point P to the point numbered i in the N known points set and z_i is the known value at point i. The exponent u is used to control the smoothness of the interpolation. As the distance between interpolation location P and the measured sample point D_i increases, the weight of that sampled point will decrease exponentially. As P approaches a data point D_i, d_i tends to zero and the i^{th} terms in both the numerator and denominator exceeds all bounds while other terms remain bounded.

Finally, the Application module contains the user defined applications such as path-finding or isopleths maps. The application module also has direct access to the In-network Processing module to get raw data if required. Figure 1 show the mapping service architecture and interaction between its four modules.

VIII. Experimental Evaluation

The efficiency of the proposed mapping service in terms of the quality of the produced map was demonstrated using the following experiment that simulates distributed execution on a real life data-set. As an example of the quality obtainable, and what might be expected from a sensor network the following maps, derived from a series presented by Clarke and Swaze [19], are presented. The maps generated by the mapping service are compared to original Fe map taken from [19].

This algorithm was implemented using a mapping API with an in-house simulation software "Dingo" [20] which is a fork of the "SenSor" project [21]. It has proven that it is not only easy to use, but also powerful enough to model and simulate the behaviour of the mapping service at various design stages. It provides an easy way to develop system models, enabling users to quickly manipulate hardware elements and achieve the desired results without having to build a full hardware prototype.

Figure 2 is derived directly from [19] and shows distribution of iron minerals around Cuprite, Nevada. This map represents an area 2km on a side, and provides a very detailed account of the mineral distribution in that area. This image has been used as the basis for a simulation of the results that might be expected from a sensor network, sensing for evidence of the same chemicals. Using the Dingo simulator, a network of sensors were randomly distributed over the 2km square, and the values of the image in Figure 3 used as the output of each sensing device at that point. The simulated sensor network was programmed to produce a map, using the Shepard interpolation method. The interpolated map of iron minerals generated by the mapping service is clearly of poorer quality than the original which could have been obtained using an orbiting image spectrometer. However, the interpolated terrain is clearly similar to the real surface. Determining exactly how close the similarity is, and what the algorithmic limits to the accuracy of the representation, is a problem we are currently investigating. Consider, though, that the information used to reconstruct the surface in Figures 2 is just 3000 points, while the original is recorded by the satellite spectroscopy which is a hard target to hit. The satellite spatial resolution is about 17 meters pixel spacing. Taking the position of the nodes into account as extra information, the reconstruction is built using less than 3% of the original data.

To highlight this potential use of an in-network mapping service, we present an implementation of isopleths generation. The results of generating isopleths based on this data are shown in Figures 4 and 5. These contours were generated in the 0.4 to 1.2 micron spectral region and a threshold of 1. Compared to the isopleths based on the actual Fe distribution map shown in Figure 4, it is visually evident that the isopleths based on the interpolated Fe distribution map, Figure 5, are visually similar to those in Figure 4.

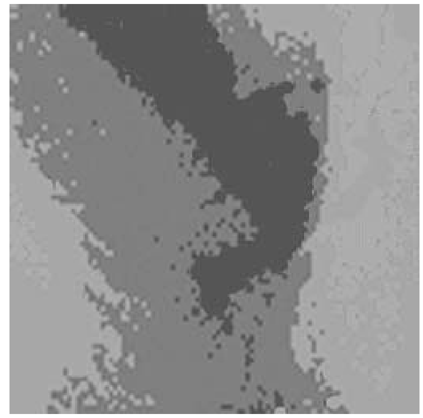

Figure 2. Fe distribution around Cuprite, NV [19].

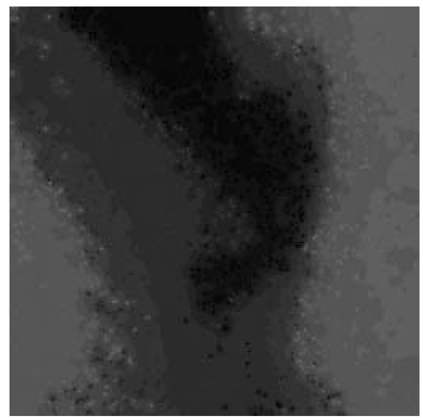

Figure 3. Interpolated map of Fe distribution around Cuprite, NV produced using 3000 sensor nodes.

IX. CONCLUSION AND FUTURE WORK

In this paper we show that visualisation of sense data gathered from networks of wireless sensors is a challenging problem in several regards. We discuss these challenges and propose a suitable data extraction and visualisation framework. We also explain why a map is a suitable data presentation format and propose an implementation of the mapping service for wireless sensor networks. We have identified applications, such as isopleths generation, that the service would make simple to develop, and challenges that need to be met before such a service is feasible. Finally, we examine the applicability of Shepard interpolation in the reconstruction of a parameter map from sparsely sampled data.

This paper is not the result of a completed project, but the exposition of the start of one. We feel that this area of research is pertinent to modern wireless sensor networks, and in this paper we have taken initial steps towards exploring it. The problems and challenges described in Section 4 are the opportunities we intend to take and the lines we intend to follow.

REFERENCES

[1] Wikipedia, "Map," 2007, [Online; accessed 26-November-2007]. [Online]. Available: http://en.wikipedia.org/wiki/Map

[2] D. Estrin, "Reflections on wireless sensing systems: From ecosystems to human systems," in *Radio and Wireless Symposium*, 2007, pp. 1 – 4.

[3] X. Meng, T. Nandagopal, L. Li, and S. Lu, "Contour maps: monitoring and diagnosis in sensor networks," *Comput. Netw.*, vol. 50, no. 15, pp. 2820–2838, 2006.

[4] R. Tynan, G. O'Hare, D. Marsh, and D. O'Kane, "Interpolation for wireless sensor network power management," in *International Workshop on Wireless and Sensor Networks (WSNET-05)*. IEEE Press, June 2005.

[5] A. D. Wood, J. A. Stankovic, and S. H. Son, "Jam: A jammed-area mapping service for sensor networks," in *RTSS '03: Proceedings of the 24th IEEE International Real-Time Systems Symposium*. Washington, DC, USA: IEEE Computer Society, 2003, p. 286.

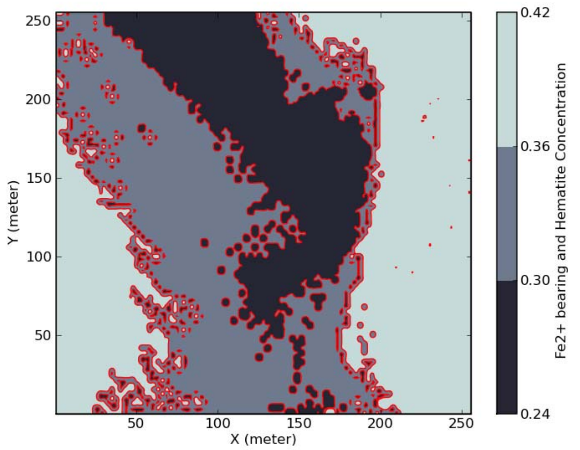

Figure 4. Isopleths added to Fe distribution map in Figure 2.

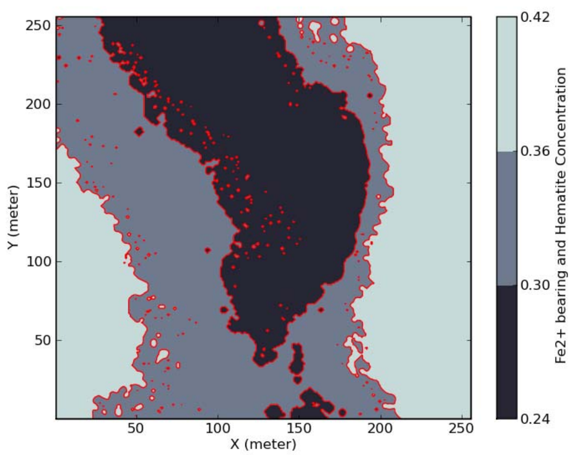

Figure 5. Isopleths added to interpolated Fe distribution map in Figure 3.

[6] W. Xue, Q. Luo, L. Chen, and Y. Liu, "Contour map matching for event detection in sensor networks," in *SIGMOD '06: Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. New York, NY, USA: ACM Press, 2006, pp. 145–156.

[7] H. Kang and X. Li, "Power-aware sensor selecion in wireless sensor networks," in *The 5th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2006.

[8] I. F. Akyildiz, M. C. Vuran, and O. B. Akan, "On exploiting spatial and temporal correlation in wireless sensor networks," *IEEE WiOpt'04, University of Cambridge*, March 2004.

[9] I. Solis and K. Obraczka, "In-network aggregation trade-offs for data collection in wireless sensor networks," INRG Technical Report 102, Tech. Rep., 2003.

[10] R. Sugihara and A. Chien, "Accuracy-aware data modeling in sensor networks," in *3rd international conference on Embedded networked sensor systems*, 2005, pp. 282–283.

[11] Z. Li, Q. Zhu, and C. Gold, *Digital Terrain Modelling: principles and methodology*, G. I. Science, Ed. CRC Press, 2005.

[12] S. Pai, "Efficient visualization of streaming sensor network data using approximation technique," http://www.cse.uta.edu/news/seminars/ShowAbstract.asp?pageVer=grph\&type=MS\&id=385, 2007, [Online; accessed 05-June-2007].

[13] B. Hamann and K. Joy, "Visualization methods for point data in space," http://www.citris-uc.org/research/projects/visualization_methods_for_point_data_in_space, 2007, [Online; accessed 05-June-2007].

[14] T. He, S. Krishnamurthy, J. A. Stankovic, T. Abdelzaher, L. Luo, R. Stoleru, T. Yan, L. Gu, J. Hui, and B. Krogh, "Energy-efficient surveillance system using wireless sensor networks," in *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*. New York, NY, USA: ACM, 2004, pp. 270–283.

[15] R. A. Becker, S. G. Eick, and A. R. Wilks, "Visualizing network data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 1, no. 1, pp. 16–28, 1995.

[16] M. Hammoudeh, A. Kurtz, and E. Gaura, "MuMHR: multi-path, multi-hop, hierarchical routing," in *International Conference on Sensor Technologies and Applications (SENSORCOMM2007)*, 2007.

[17] W. Heinzelman, A. Chandrakasan, and H. Balakrishna, "Energy-efficient communication protocol for wireless microsensor networks," *Proceedings of the 33rd International Conference on System Sciences*, January 2000.

[18] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM national conference*. ACM Press, 1968, pp. 517–524.

[19] Clark and Swayze, "Summaries of the 6th annual jpl airborne earth science workshop," in *JPL Airborne Earth Science*, March 1996, pp. 4–8.

[20] S. Mount, "Dingo wireless sensor networks simulator," http://code.google.com/p/dingo-wsn/, 2008, [Online; accessed 26-October-2008].

[21] S. Mount, R. Newman, E. Gaura, and J. Kemp, "Sensor: an algorithmic simulator for wireless sensor networks," in *In Proceedings of Eurosensors 20*, vol. II, Gothenburg, Sweden, 2006, pp. 400–411.